Pinocchio: Conducting a Virtual Symphony Orchestra

Bernd Bruegge, Christoph Teschner, Peter Lachenmaier, Eva Fenzl,
Dominik Schmidt, Simon Bierbaum
Chair for Applied Software Engineering
TU München
85748 Garching, Germany
{bruegge, teschner, lachenma, fenzl, schmidtd, bierbaum}@in.tum.de

Abstract
We present a system that allows users of any skill to conduct a virtual orchestra. Tempo and volume of the orchestra's performance are influenced with a baton. Pinocchio works with several types of batons, differing in tracking method and in algorithms for gesture recognition. The virtual orchestra can be configured, allowing the muting, hiding and positioning of individual musicians or instrument groups in 3D space. The audio and video material is based on a professional recording session with the Bavarian symphony orchestra. Pinocchio's long-term goal is the creation of a multi-modal, device independent framework for gesture-based applications which require motor skills or the control and operation of a complex set of sensors in intelligent house or car driver assistance systems. In this paper, we describe the current development status of the project, detail its usage and finally give an overview over our future project goals.

Categories and Subject Descriptors
J.5 [**Arts and Humanities**]: Performing Arts

General Terms
Algorithms, Performance, Design, Experimentation, Human Factors

Keywords
Virtual Symphony Orchestra, classical music, conducting, gesture recognition, 3D audio, 3D video, neural networks, Wii Remote, eWatch.

1. INTRODUCTION
Pinocchio is an example of a class of currently emerging systems that are able to understand and react to the *intuition* of a user. We define intuition as a way of experiencing objects (opposed to opinion based on that experience). The goal of the Pinocchio project is to enable children to experience and develop an understanding for classical music in a playful manner by conducting a virtual orchestra. We are developing Pinocchio as a game to be usable in a living room environment as well as in a museum setting.

The user is able to configure the orchestra, which is modeled as a set of individually controllable musicians. Each musician is associated with audio and video streams and can be freely located in 3D space.

The orchestra adjusts to the tempo and volume according to the conductor's gestures. Thus the system must be capable of performing synchronous real-time video and audio time stretching. Pinocchio supports tracking the conductor's eyes. Combining the eye movements with information from the score, a context is established to interpret the gesture as a command for an individual musician or a group within the orchestra. The interaction must be intuitive and easy to learn because children and amateurs are the targeted user group for the system. The system supports different input devices such as regular conductor batons as well as devices measuring the movement of the wrist or hand.

ACE'07, June 13-15, 2007, Salzburg, Austria.
ACM 978-1-59593-640-0/07/0006.

2. DEMONSTRATION
Since we develop Pinocchio to be used in a typical living room, our demonstration is easy to setup - it consists of a laptop with a built-in camera, headphones, a regular conductor's baton and Nintendo's Wii Remote.

The virtual orchestra is freely configurable regarding presence and position of individual musicians in virtual 3D space. Pinocchio has a constantly visible control panel which can be used to set a musician's coordinates and to mute each musician. Muted musicians become invisible and inaudible, enabling the conductor to listen for instance to a single cello if he wishes to.

Users can choose from two different input devices: A regular conductor's baton and the Wii Remote.

The baton is tracked via camera. To facilitate tracking and to improve user experience, a small colored ball is attached to the baton's tip. To begin conducting, the user picks up the baton and holds it so that it is within the camera's viewing frustum. Pinocchio features a small control window on screen which shows the camera view augmented by a trace of the baton tracking and the currently recognized conduction gesture. The virtual orchestra starts its performance when it receives a start command, issued by either a special baton gesture or the start button on Pinocchio's on-screen control panel. The orchestra changes tempo and volume according to the conductor's gestures.

Nintendo's Wii Remote is a commercial product built as a wireless controller for the Wii gaming console. It features three acceleration sensors as well as a set of buttons on the top surface, packaged into an enclosure with a remote control form factor. Apart from pairing with the computer, we do not use the buttons but rely solely on input through the accelerometers. Pinocchio recognizes a start and a stop gesture and extracts tempo and volume information from an easy-to-learn conducting motion.

Besides getting quality feedback, we pursue one specific goal with the demonstration: observing many users of Pinocchio at an open day at our university, we found that their gestures, while looking very similar, created distinctly different acceleration patterns and confused our early gesture recognition system. We would like to take the opportunity to record and analyze conducting sessions with conference participants.

3. RELATED WORK

The first system that provided an interactive conducting experience was Mathews' Radio Baton [9]. The Digital Baton [10] measures the position of the baton along with additional parameters, such as pressure on the handle in order to obtain more precise input. The Conductor's Jacket [11] uses additional sensors for tracking muscle tension. The MultiModal Conducting Simulator [12] uses Hidden Markov Models for gesture recognition. Ilmonen's Virtual Orchestra [13] features MIDI-based audio and synthetically rendered graphical output. Murphy, Andersen and Jensen first use a real baton tracked with a camera [14]. iSymphony [7] is a conducting system developed by the Media Computing Group at RWTH Aachen University, Germany. It is a mature system and the latest incarnation of the Personal Orchestra [8]. "EyeToy" is a commercially available gesture-controlled game designed for Sony's PlayStation 2. "Maestro" is a special EyeToy game targeted for conducting.

4. SOFTWARE ARCHITECTURE

Pinocchio consists of three main subsystems responsible for video playback, audio playback and gesture recognition.

The audio subsystem allows the manipulation of multi-channel audio data in various ways. Its core functionality includes mixing audio streams from several different sources and then placing it in 3-dimensional space. Further, the subsystem allows dynamic tempo modification in real-time without affecting pitch, as well as modifications of dynamics and various reverb-preset settings.

The video subsystem is responsible for displaying the orchestra and the environment in which it plays. Individual musicians can be moved around freely in their surrounding, they can be hidden and the user can see the orchestra from an arbitrary point of view. To create a believable setting, a concert hall is rendered in the background.

The tracking and gesture recognition subsystem analyzes the conductor's hand movement. Pinocchio can currently switch between two methods. The first method is video based baton tracking and gesture recognition using neural networks. The second method is acceleration-based. We experimented with two different hardware controllers, the eWatch [4] and the Wii Remote [3]. They measure acceleration and transmit data via Bluetooth to the connected computer. With both devices we can use a fast Fourier transform (FFT) to determine tempo and volume.

5. FUTURE WORK

We plan to work on a multitude of topics:

- Scaling: We are working on optimizing the application to accommodate up to 90 musicians.
- Score integration: We would like to make the score of the piece easily accessible from within Pinocchio. The user should be able to interact with it, for instance to start the performance at an arbitrary point.
- Machine learning: More gesture recognition algorithms will be evaluated in order to improve user experience. Accuracy and latency should be improved.
- Input methods and data fusion: Using new and concurrent input data such as voice or face tracking, Pinocchio will become more capable of finding out and acting according to the user's intention.

A long-term goal is to create a generalized framework for multi-modal, context-sensitive user interaction based on experience gained from Pinocchio.

6. ACKNOWLEDGEMENTS

Our thanks go to the Bavarian Radio Symphony Orchestra for providing valuable assistance and the audio and video recordings. We also want to thank the Carnegie Mellon University for giving us the chance to work with the eWatch.

7. REFERENCES

[1] *Symphonieorchester des bayerischen Rundfunks*, http://www.br-online.de/kultur-szene/klassik/pages/so/, 2007

[2] *Virtual Symphony Orchestra*, http://wwwbruegge.in.tum.de/VSO, 2006

[3] *Wii Remote*, http://wii.nintendo.com/controller.jsp, 2007

[4] Maurer, U., Smailagic, A., Siewiorik, D. *eWatch: A Wearable Sensor and Notification Platform*. In *Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks (BSN '06)* (MIT, Cambridge, USA, April 3-5, 2006). IEEE Conference Proceedings, Los Alamitos, USA, 2006, 4 pp.

[5] *DarwiinRemote*, http://sourceforge.net/projects/darwiin-remote/, 2007

[6] Lachenmaier, P., *Entwurf und Implementierung eines Dirigierlernsystems für Kinder*, bachelor's thesis at Technische Universität München, 2006

[7] Lee, E., Karrer, T., Kiel, H., et al, *iSymphony: An Adaptive Interactive Orchestral Conducting System for Digital Audio and Video Streams.* In Proceedings of CHI 2006 (Montreal, Canada, April 22-27, 2006). 259-262

[8] *Personal Orchestra*, http://media.informatik.rwth-aachen.de/po.html, 2006

[9] Mathews, M. V., *Current Directions in Computer Music Research,* chapter The Conductor Program and Mechanical Baton. MIT Press, Cambridge, 1991

[10] Marrin, T., *Possibilities for the digital baton as a general-purpose gestural interface.* CHI, pages 311-312. ACM, 1997

[11] Marrin, T., *Inside the Conductor's Jacket: Analysis, interpretation and musical synthesis of expressive gesture.* PhD thesis, Massachusetts Institute of Technology, 2000

[12] Satoshi, U., Yasunori, M. *A conducting recognition system on the model of musicians' process*. Journal of the Acoustical Society of Japan, 1998

[13] Ilmonen, T., *The virtual orchestra performance.* CHI. ACM, 2000

[14] Murphy, D, Andersen, T. H., *Jensen, K. Conducting audio files via computer vision*. Proceedings of the Gesture Workshop, Genova, 2003